

# Analysis of multienvironment trials (MET) in the sugarcane breeding program of Réunion Island

Susie Guilly · Thomas Dumont · Audrey Thong-Chane · Laurent Barau · Jean-Yves Hoarau

Received: 4 January 2017 / Accepted: 7 August 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** The sugarcane variety development program on Réunion Island is dedicated to an industry that encompasses numerous different agroclimatic production zones. The objective of this study is to characterize in detail the final selection stage of this program, consisting of multienvironment trials (MET) at seven representative locations, considering the genotypic response in terms of tonnes of cane per hectare (TCH), estimable recoverable sugar (ERS), fiber content (FIB), and an economic index (EI). Data from four recent variety series tested between one and three crop years were used. Each trait revealed a significant genotype  $\times$  location (GL) interaction, always higher than the genotype  $\times$  crop year (GC) interaction, indicating that testing genotypes across locations is more important than testing for ratooning ability. Broad-sense heritability ( $H$ ) at MET level was relatively high for FIB compared with EI, while ERS and TCH exhibited intermediate  $H$  values. Genotype main

effect plus genotype  $\times$  environment (GGE) biplot analysis applied to a balanced set of genotypes tested over two crop years across all environments permitted reliable visualization at a glance of (i) the level of proximity between genotypes or environments, and (ii) the performance of varieties in each environment and their stability across all of them. No redundancy between any pair of environments was found for the most important selection trait (EI). These results confirm the relevance of a selection strategy firstly oriented toward selecting sugarcane genotypes for local adaptations with the objective of enhancing the mean productivity of the whole cane industry.

**Keywords** Sugarcane · Multienvironment trials (MET) · Genotype  $\times$  location (GL) interaction · Genotype main effect plus genotype  $\times$  environment (GGE) biplot

---

Susie Guilly and Thomas Dumont contributed equally to this paper.

---

S. Guilly · T. Dumont · A. Thong-Chane · L. Barau · J.-Y. Hoarau (✉)  
eRcane, 29 rue d'Emmerez de Charmoy,  
97490 Sainte-Clotilde, La Réunion, France  
e-mail: hoarau.jean-yves@ercane.re

J.-Y. Hoarau  
Cirad, UMR AGAP, Avenue Agropolis,  
34398 Montpellier, France

## Introduction

Sugarcane (*Saccharum* spp.) is the world's most important biomass-producing crop. It is a cash crop grown over semiperennial durations, with harvesting occurring at yearly intervals from time of planting. Breeding programs for the cane industry aim to select new genotypes with local or broad adaptation to provide high yields over several crop years within target environments representative of different

cultivation areas. Breeding programs consist of several successive stages of selection using clonal multiplication. In the most advanced selection stages, the best elite candidates are tested in multi-environment trials (MET) at representative locations using replicated experiments. MET carried out over several crop years allow testing of genotype  $\times$  location (GL) and genotype  $\times$  crop year (GC) interactions. These two components of the genotype  $\times$  environment (GE) interaction determine the yield potential and yield stability of sugarcane varieties. They have been quantified for cane yield and sucrose content or for derived sucrose yield in breeding programs at various locations, including Australia (Jackson and Hogarth 1992; Mirzawan et al. 1994), Florida (Milligan et al. 1996; Glaz and Kang 2008), South Africa (Parfitt 2000; Ramburan et al. 2012a), Venezuela (Rea and De Souza Vieira 2002), and Argentina (Mariotti and Clariana 1994). All these examples of MET studies encompassed locations spread over relatively large cultivated areas, necessarily made up of environments that are more or less heterogeneous in various regards, such as soil type (Milligan et al. 1996; Glaz and Kang 2008) or cultivation practices (Mirzawan et al. 1994). Studies providing formal statistical tests of interactions revealed that both GL and GC effects were always significant for all traits surveyed, except on one occasion (Parfitt 2000). Moreover, GL interactions were more important than GC interactions in these reports.

G  $\times$  E interactions may cause changes in the relative ranking of genotypes across sites and/or crop years in trials and complicate identification of superior cultivars by confounding determination of true genetic values. When G  $\times$  E interactions exist, their statistical significance and precise characteristics must be investigated in detail to assess implications for selection strategies and help optimize resource allocation across locations and years. As a first step, analysis of variance of MET data provides a general picture of the influence of the different factors underlying phenotypic variation. In a second step, genotype plus genotype  $\times$  environment interaction (GGE) biplot analysis (Yan and Tinker 2005, 2006) is a very useful graphical tool to investigate in detail the relationships between environments and the pattern of the response of genotypes across environments. This popular visualization technique for MET data has been used

in several sugarcane programs to (i) investigate the similarity of environments and their ability to discriminate genotypes (Glaz and Kang 2008; Ramburan et al. 2012a; Luo et al. 2015), (ii) identify redundant sites or mega-environments and analyze the stability of the site response across series of genotypes (Ramburan et al. 2012a, b), and (iii) visualize the performance rank and stability of genotypes across environments for the purposes of decision-making regarding release of new cultivars (Glaz and Kang 2008; Shandu et al. 2012; Klomsa-ard et al. 2013; Luo et al. 2015).

On Réunion Island, production of sugarcane is scattered over many different agroclimatic regions on both leeward (dry) and windward (wet) coasts. These different regions range in altitude from sea level to highlands up to 900 m and consist of vastly different soil types in terms of physical properties and chemical fertility. Commercial sugarcane genotypes are selected and released by eRcane research institute. eRcane currently operates seven selection programs (eRcane 2009), strategically located in the major sugarcane growing areas (Table 1). Programs run concurrently to identify promising genotypes suited for each zone and currently take 14 years from time of initial cross to new cultivar release. During the last four testing years, these seven programs have shared a common final selection stage, gathering a set of elite genotypes selected from each of the seven selection sites. Each year, a new MET series of genotypes is planted across the seven locations and tested over several crop years. Since the inception of the present-day MET network in the late 2000s (eRcane 2009), no study has been conducted to evaluate the GE interactions and the relationships among the locations in terms of genotypic response.

The present study aims to investigate the characteristics of the present-day MET selection program on Réunion Island by retrospective analysis of some recent genotype series tested across the whole network of selection stations. The objectives are to: (i) assess variance components of the major quantitative traits relative to yield components and estimate the importance of GE interactions, (ii) study relationships existing between environments in terms of genotypic response, (iii) test the added value that the GGE biplot statistical tool can bring to provide support in decision-making for selection, and (iv) identify potential areas for optimizing the current MET program.

**Table 1** Climate and soil description of the seven eRcane locations used to test advanced MET selection series

Geographic zone	Windward coast			Leeward coast		
	LM: La Mare	SB: Saint-Benoit	MN: Mencil	ES: Étang-Salé	GL: Le Gol	VB: Vue-Belle
Location						SP: Saint-Philippe
Latitude	20.91°S	21.06°S	20.98°S	21.26°S	21.27°S	21.08°S
Longitude	55.53°E	55.72°E	55.60°E	55.37°E	55.40°E	55.29°E
Focus of selection	Pluvial coastal area with water deficit	Pluvial area, favorable rainfall	Over-humid high area	Dry coastal area, with irrigation supply	Dry coastal area, with hydric stress	Highland with water deficit
Climatic parameters						
Altitude above sea level (m)	21	40	400	20	20	700
Annual average rainfall (mm)	1470	3290	3250	614	614	870
Temperature range (°C)	20.9–26.4	20.4–26	19.3–25.6	20.3–27.3	20.3–27.3	14.5–21.5
Number of months with water deficit (evapotranspiration higher than rain)	8/12	0/12	0/12	12/12	12/12	7/12
Average relative humidity (%)	70	75–80	80	70	70	75
Soil type						
Classification	Ferrallitic	Andic brown	Andosol	Fine alluvium	Fine alluvium	Andic brown
Main physical characteristics	Deep, well structured, few stones	Very stony, heterogeneous	Shallow with many stones	Deep, loose, slightly hydromorphic	Deep, loose, slightly hydromorphic	Good structural stability, highly permeable
pH	Medium	Medium	Low	Relatively high	Relatively high	Low
Cation exchange capacity	Low	Low	Low	High	High	Very low
Chemical fertility	Satisfactory	Satisfactory	Low	High	High	Low
						Modest
						Scoriaceous lava
						Very stony, heterogeneous, high microporosity level

## Materials and methods

### Multienvironment trial (MET) dataset

The MET analyzed in this study were carried out at a network of seven selection sites on Réunion Island, France: La Mare (LM), Saint-Benoit (SB), Menciol (MN), Étang-Salé (ES), Le Gol (GL), Vue-Belle (VB), and Saint-Philippe (SP). These sites cover a wide range of ecologies of production representative of the main sugarcane growing areas. The main climatic and soil type characteristics of these seven locations are summarized in Table 1. Sites represent either coastal zones (LM, SB, ES, and GL) or zones at low (SP) or medium (MN) altitude, as well as highland (VB) characterized by cooler temperatures. The variability in terms of soil type and physical characteristics for cane cultivation is relatively large. Chemical fertility is also highly variable, being either favorable (ES, GL), satisfactory (LM, SB, SP), or very low (VB, MN).

The MET dataset consists of four consecutive genotype series (S00, S03, S04, and S05) planted in the years 2011–2014 at the seven sites. Each MET series was tested in a randomized complete block design with four replications at all locations. Each plot had area of 45 m<sup>2</sup> with three rows, with length of 10 m and width of 4.5 m. As indicated in Table 2, MET series were phenotyped and harvested at all sites for between one and three crop years, according to the year of their planting. Due to constraints in terms of land resources and the high cost of the whole breeding

program, the material composing a final MET series usually represents a set of genotypes that is not balanced across all sites. On average, only about half of the clones tested at one location were also tested at the other six locations. Altogether, the four MET series studied herein represent a balanced number of 47 genotypes tested across all sites. Depending on the site considered, two (LM, MN, SP, VB, SB), three (ES), or four (GL) of the best local cultivars were also planted as common standards in the trials of the four genotype series (see footnote b of Table 2).

At the end of each crop year, four quantitative traits were recorded at individual plot level: tonnes of cane per hectare (TCH, Mg ha<sup>-1</sup>), fiber content as percentage of fresh weight (FIB, %) estimable recoverable sugar (ERS, %) and an economic index (EI). All millable stalks from each variety plot were manually cut and weighed using a digital scale mounted on a tractor-operated hydraulic boom. TCH was determined from plot weights divided by plot areas. From each plot, a sample of 18 randomly selected stalks was used to determine ERS and FIB at eRcane laboratory using the standard hydraulic press method (Hoarau 1969): FIB was determined from the weight of the press cake from a 500-g shredded subsample. ERS was calculated from the FIB value and Brix (Bellingham RFM340 refractometer) and Pol (Polaser SR 64 polarimeter) measurements of the extracted juice according to conventional calculations used in the local cane industry adapted from Saranin (1986). The conventional economic index (EI) used by eRcane to rank the cultivation merit of candidate genotypes was

**Table 2** Summary of the four advanced MET selection series (S00, S03, S04, and S05) tested at the seven locations described in Table 1

Series	Year planted	Number of crop years studied	Number of genotypes investigated at each location <sup>a</sup>							Number of genotypes common to all locations
			LM <sup>b</sup>	SB <sup>b</sup>	MN <sup>b</sup>	ES <sup>b</sup>	GL <sup>b</sup>	VB <sup>b</sup>	SP <sup>b</sup>	
S00	2011	3	18	16	17	19	22	19	18	10
S03	2012	2	21	21	17	17	21	13	19	11
S04	2013	1	21	21	22	21	22	18	19	11
S05	2014	1	22	21	24	21	22	17	21	15
Total										47

<sup>a</sup> Locations: LM La Mare, SB Saint-Benoit, MN Menciol, ES Étang-Salé, GL Le Gol, VB Vue-Belle, SP Saint-Philippe

<sup>b</sup> Cultivar checks added in all four series trials at each location: R579 and R585 at LM, MN, and SP; R577 and R583 at VB; R579 and R582 at SB; R579, R582, and R584 at ES; R570, R579, R584, and R585 at GL

calculated from the TCH and ERS using the formula  $EI = TCH \times (ERS - 4)$ , being indicative of farming profit net of production costs (Hugot 1958).

### Variance components analysis

Variance components for each trait were assessed using mixed linear models designed to estimate GL and GC interaction effects. This was first done separately for each MET series consisting of unbalanced numbers of genotypes among sites, with the following mixed linear model fit to each trait:

$$\begin{aligned} Y_{jklm} = & \mu + L_j + C_{k(j)} + R_{l(j)} + \underline{G}_m + \underline{GL}_{mj} + \underline{GC}_{mk} \\ & + \underline{GLC}_{mjk} + e_{jklm} \end{aligned} \quad (\text{model 1}),$$

where  $Y_{jklm}$  is the observation of genotype  $m$  in replicate  $l$  in crop year  $k$  at location  $j$ ,  $\mu$  is the grand mean,  $L_j$  is the location main effect at  $j$ ,  $C_{k(j)}$  is the crop year effect for  $k$  at location  $j$ ,  $R_{l(j)}$  is the replication effect for  $l$  at location  $j$ ,  $G_m$  is the main effect for genotype  $m$ ,  $GL_{mj}$  is the effect of the genotype  $\times$  location interaction,  $GC_{mk}$  is the effect of the genotype  $\times$  crop year interaction,  $GLC_{mjk}$  is the effect of the genotype  $\times$  crop year  $\times$  location interaction, and  $e_{jklm}$  is the error. Effects considered as random in the model are underlined, while others were fixed. In the particular cases of series S04 and S05, which were studied in a single year, the crop year effect ( $C_{k(j)}$ ) was dropped from the model.

Combined global analysis of all four MET series was carried out using the subset of 47 genotypes that were commonly tested at all seven sites using the following mixed linear model fit to each trait:

$$\begin{aligned} Y_{ijklm} = & \mu + S_i + L_j + C_{k(i,j)} + R_{l(i,j)} + \underline{G}_m + \underline{GL}_{mj} \\ & + \underline{GC}_{mk} + \underline{GLC}_{mjk} + e_{ijklm} \end{aligned} \quad (\text{model 2}),$$

where  $S_i$  is the main effect of series  $i$ , while the other terms and indices are the same as in model 1.

All models were performed using the MIXED procedure of SAS 9.3 (SAS Institute, Cary, NC), and variance component estimates were evaluated using the restricted maximum likelihood (REML) procedure. Use of the COVTEST option in the model statement provided standard errors of variance components and a derived Wald Z-test of their statistical

significance (Littell et al. 2006). In both models 1 and 2, measurements acquired on the same individual plot across successive crop years in series S00 and S03 were considered as repeated measurements (longitudinal data).

When analyzing series S00 and S03 with model 1, four **R** matrices of the variance–covariance (VCV) of random error terms were tested to select the structure that modeled the data optimally: (i) using the UN option to allow all VCV parameters to be unstructured (pairs of within-genotypes errors having their own correlations), (ii) using the UN(1) option to specify the constraint of covariances of error terms between crop years to be null, (iii) using the first-order autoregressive AR(1) option with the constraint of correlation between adjacent within-genotypes errors  $|\rho| < 1$ , and (iv) using a component symmetry (CS) option having the constraint of constant variances and constant covariances between error terms. The goodness of fit values for the UN, UN(1), AR(1), and CS models were compared using the Akaike information criterion (AIC) (Akaike 1974), which balances model fit versus number of parameters.

When analyzing all four MET series together, an UN structure was considered for the **R** matrix.

Broad-sense heritability  $H$  or genetic repeatability refers to the extent to which the phenotype is determined by its genotype (Falconer and Mackay 1996). At MET level,  $H$  was calculated for each trait with variance components estimated from each model, as a ratio of the genetic variance ( $\sigma_G^2$ ) to phenotypic variance of genotype means ( $\sigma_p^2$ ):

$$H = \frac{\sigma_G^2}{\sigma_p^2},$$

where  $\sigma_p^2 = \sigma_G^2 + \frac{\sigma_{GL}^2}{L} + \frac{\sigma_{GC}^2}{C} + \frac{\sigma_{GLC}^2}{LC} + \frac{\sigma_e^2}{LCR}$ , with  $\sigma_{GL}^2$ ,  $\sigma_{GC}^2$ , and  $\sigma_{GLC}^2$  designating GL, GC, and GLC interaction variances, respectively, and considering  $L = 7$  locations,  $R = 4$  replicates in both models, with  $C = 1\text{--}3$  crop years in model 1 depending on the MET series considered and  $C$  equal to the harmonic mean of the number of crop years of the different series in model 2 (Holland et al. 2003, pp. 64–65).

Model 2, which was used to analyze the combined MET series, was considered to provide more representative estimates of variance components than those obtained from model 1, because of the higher number of genotypes used to compute the estimates (47 instead

of about 20) and the subsequent smaller risk of high individual influence of genotypes on variance component partitioning. Using the variance component estimates obtained from model 2, we simulated the broad-sense heritability ( $H$ ) of traits at MET level using various combinations of number of locations ( $L$ ) and crop years ( $C$ ) up to the maximum numbers in the experiments ( $L = 7$  and  $C = 3$ ) and considering a standard trial design with  $R = 4$  replications. The changes in  $H$  when the number of locations and replications were increased were plotted graphically to explore the effect of these changes and assess the influence of experimental effort on the accuracy of the mean performance of genotypes across all sites.

### GGE biplots

To visualize the relationships existing among selection sites and the performance of some of the most promising candidate genotypes, GGE biplots were produced for the 21 genotypes of both series S00 and S03 (10 and 11 genotypes, respectively) that were tested across all sites (Table 2). To this end, we restricted the trait data analyzed to the first two crop years in order to consider space–time information perfectly balanced (7 locations  $\times$  4 replications  $\times$  2 crop years  $\times$  21 genotypes) to visualize relationships among environments and genotypes without risks of bias. In a first step, using cultivar checks common to both series within each site (see footnote b of Table 2), both series were analyzed at each site using the following linear mixed model:

$$Y_{iklm} = \mu + S_i + C_{k(i)} + R_{l(i)} + G_m + e_{iklm} \quad (\text{model 3}),$$

where the symbols and indices designate the same effects as in model 2 and in which the genotype effect was fixed and other effects were random. At each location  $j$ , the adjusted mean ( $\bar{Y}_{mj}$ ) of genotype  $m$  (free of any trial/series membership effects) was inferred by adding to the best linear unbiased estimate (BLUE) of genotype the overall mean ( $\mu$ ) of the model (Littell et al. 2006, p. 211). Pearson's correlation coefficients between traits and across sites were calculated using average adjusted means of genotypes for the seven sites. For each trait, the contribution of each genotype  $m$  to the genotype  $\times$  location interaction was estimated using Wricke's (1965) ecovalence stability

index ( $W_m$ ), and the contribution of each location  $j$  to the interaction was estimated using a similar index ( $W_j$ ) by interchanging locations and genotypes as follows:

$$W_m = \sum_{j=1}^7 (\bar{Y}_{mj} - \bar{Y}_{m.} - \bar{Y}_j - \bar{Y}_{..})^2,$$

$$W_j = \sum_{m=1}^{21} (\bar{Y}_{mj} - \bar{Y}_{m.} - \bar{Y}_j - \bar{Y}_{..})^2,$$

where  $\bar{Y}_{m.}$ ,  $\bar{Y}_j$ , and  $\bar{Y}_{..}$  are the adjusted means of genotypes and locations, and overall mean, respectively. The two-way data table of adjusted genotype means  $\times$  locations ( $\bar{Y}_{mj}$ ) was then centered to the mean trait value ( $\mu_j$ ) of each environment  $j$  and divided by its standard deviation ( $s_j$ ) to obtained a “standardized GGE matrix” of the genotype main effect (G) and genotype  $\times$  environment interaction (GE). This standardized GGE matrix was subjected to singular value (SV) partitioning between the genotype and environment eigenvectors using the general models of Yan (2002) and Yan and Tinker (2005, 2006):

$$(\bar{Y}_{mj} - \mu_j)/s_j = \sum_{p=1}^n \lambda_p^{f_p} \alpha_{mp} \lambda_p^{1-f_p} \gamma_{jp} + e_{mj},$$

where  $\lambda_p$  is the SV of the  $p$ th principal component (PC),  $\alpha_{mp}$  and  $\gamma_{jp}$  are the respective eigenvectors of genotype  $m$  and environment  $j$  for PC  $p$ ,  $e_{mj}$  is the residual associated with genotype  $m$  in environment  $j$ , and  $f_p$  is a partition factor for PC  $p$ , equal to 0.5 for all PCs  $p$ , to visualize genotype scores and environmental scores in the same units for both PC1 and PC2 (Yan 2002).

GGE biplots provide, at a glance, the ranking of all genotypes regarding their performance in any environment, when visualizing the distribution of their positions when projected orthogonally onto each environment axis (Yan and Tinker 2006). To assess the efficiency of our GGE biplots for visualizing the trait performance of genotypes in each environment, we calculated for each environment a coefficient of correlation between the rank of genotypes in the original data and their apparent rank along the environment axis. The latter was precisely inferred by computing genotype abscissa on the axis of each environment  $j$  resulting from the scalar product between the vector of each genotype  $m$  ( $\bar{u}_m$ ) and the



environment vector ( $\vec{v}_m$ ) divided by the length of the latter ( $\vec{u}_m \cdot \vec{v}_m / \|\vec{v}_m\|$ ) (Spiegel et al. 2009).

## Results

### Variance components analysis

The AIC values for model 1 associated with the different **R** structures tested in trait analyses for series S00 and S03 are presented in Table 3. In both series, the TCH and EI data were best modeled using the UN structure and the FIB data when using the UN(1) structure (lowest AIC values), while the ERS data were best modeled when using either the UN (S00) or UN(1) (S03) structure.

Accordingly, Table 4 presents the estimates of variance components obtained from the separate analyses of the four series (model 1) or from the “global” analysis performed on a combination of balanced subsets of genotypes from each series across all sites (model 2). As a general rule, the statistical significance of genotype ( $\sigma_G^2$ ) and genotype  $\times$  location ( $\sigma_{GL}^2$ ) variance components may fluctuate for some traits across the four individual series (S00, S03, S04, and S05), as well as the genotype  $\times$  crop year ( $\sigma_{GC}^2$ ) and genotype  $\times$  location  $\times$  crop year ( $\sigma_{GLC}^2$ ) variance components across the two oldest series (S00 and S03). However, estimates of variance components obtained from all these individual analyses (model 1) were convergent with estimates obtained from the “global” model (model 2). In this reference second model, the  $\sigma_G^2$  and  $\sigma_{GL}^2$  variance components were always highly significant ( $P < 0.01$  or  $P < 0.001$ ) for

all traits, as well as the  $\sigma_{GLC}^2$  component, except for FIB since this component was estimated to be null. For  $\sigma_{GC}^2$ , this component was more or less significant for TCH ( $P < 0.01$ ) as well as for FIB and EI ( $P < 0.05$ ), but was not significant for ERS ( $P > 0.05$ ). Apart from residual variance, the order of importance of the variance components in terms of percentage of phenotypic variance explained was  $G > GL > GLC > GC$  for TCH and ERS,  $G > GL > GC > GLC$  for FIB, and  $GL > G > GLC > GC$  for EI. The genetic variance ( $\sigma_G^2$ ) represented 31, 24, 23, and only 16%, for FIB, ERS, TCH, and EI, respectively.

### Broad-sense heritability with different locations $\times$ crop years entry-mean bases

Simulated broad-sense heritability ( $H$ ) at the maximal experimental level of seven locations and three years reached 0.76 for EI, 0.85 for TCH, 0.88 for ERS, and 0.92 for FIB (Fig. 1). Accordingly, lower entry-mean heritability ( $H$ ) simulated with fewer locations and years ( $L \leq 7$  and  $C \leq 3$ ) always showed the same ranking of  $H$  values among traits, i.e.,  $EI \ll TCH \sim ERS \ll FIB$ . When considering a single location and single crop year ( $L = C = 1$ ),  $H$  was very low for EI (0.26), modest for TCH and ERS (0.38 or 0.41), and relatively high for FIB (0.55). An increase of one or two unit(s) of either  $L$  or  $C$  (or both) rapidly enhanced the  $H$  values of all traits. For  $L + C \geq 3$ ,  $H$  evolved more slowly toward final plateaus that conserved the initial ranking between traits regardless of which entry-mean basis was examined. The most predictable traits (FIB, ERS, and TCH) required less  $L + C$  to attain reliable  $H$  estimates compared with the

**Table 3** Number of fit parameters and Akaike information criteria (AIC) of different **R** matrices for model 1 applied to S00 and S03 MET data for tonnes of cane per hectare (TCH), estimable recoverable sugar (ERS), fiber content (FIB), and economic index (EI)

Trait	<b>R</b> matrix	Number of parameters	TCH <sup>a</sup>	ERS <sup>a</sup>	FIB <sup>a</sup>	EI <sup>a</sup>
S03	AR(1)	5	8854.8	2383.4	3163.0	4120.2
	CS	5	8854.8	2383.4	3163.0	4120.2
	UN(1)	5	9006.2	<b>2381.9</b>	<b>3114.3</b>	4193.0
	UN	6	<b>8839.5</b>	2383.9	3115.8	<b>4102.3</b>
S00	AR(1)	5	13,603.7	3943.1	4461.0	6232.6
	CS	5	13,534.4	3944.1	4453.2	6212.4
	UN(1)	7	13,764.9	<b>3891.4</b>	4401.7	6297.4
	UN	13	<b>13,510.1</b>	3892.4	<b>4387.9</b>	<b>6206.0</b>

<sup>a</sup> Bold figures indicate lowest AIC values

**Table 4** Estimates of variance components, percentage of phenotypic variance explained, and broad-sense heritability at experimental level for tonnes of cane per hectare (TCH), estimable recoverable sugar (ERS), fiber content (FIB), and economic index (EI) from individual (model 1) or partially global (model 2) analyses of MET series data

MET series	Model 1								Model 2	
	S05		S04		S03		S00		All series	
TCH (Mg ha <sup>-1</sup> )										
$\sigma_G^2$	160.52**	23%	153.04**	23%	151.52**	23%	181.45**	25%	166.19***	23%
$\sigma_{GL}^2$	108.78***	16%	209.39***	31%	136.69***	20%	107.07***	15%	111.37***	15%
$\sigma_{GC}^2$					34.10**	5%	12.85*	2%	19.41**	3%
$\sigma_{GLC}^2$					26.93***	4%	37.71***	5%	38.50***	5%
$\sigma_e^2$	426.80	61%	302.92	46%	323.98	48%	388.90	53%	394.27	54%
$H$	0.84		0.79		0.77		0.87		0.79	
ERS (%)										
$\sigma_G^2$	0.170**	17%	0.289**	30%	0.280**	32%	0.327**	28%	0.218***	24%
$\sigma_{GL}^2$	0.054 <sup>NS</sup>	5%	0.184***	19%	0.128***	15%	0.226***	19%	0.139***	16%
$\sigma_{GC}^2$					0.013 <sup>NS</sup>	1%	0.003 <sup>NS</sup>	0%	0.008 <sup>NS</sup>	1%
$\sigma_{GLC}^2$					0.017 <sup>NS</sup>	2%	0.079***	7%	0.047**	5%
$\sigma_e^2$	0.768	77%	0.485	51%	0.435	50%	0.549	46%	0.487	54%
$H$	0.83		0.87		0.89		0.88		0.84	
FIB (%)										
$\sigma_G^2$	0.483**	38%	0.756***	34%	1.735***	60%	0.815***	44%	0.534***	31%
$\sigma_{GL}^2$	0.126***	10%	0.084 <sup>NS</sup>	4%	0.140***	5%	0.159***	9%	0.142***	8%
$\sigma_{GC}^2$					0.058*	2%	0.014 <sup>NS</sup>	1%	0.459*	3%
$\sigma_{GLC}^2$					0.000 <sup>NS</sup>	0%	0.048*	3%	0.000 <sup>NS</sup>	0%
$\sigma_e^2$	0.657	52%	1.352	62%	0.993	34%	0.830	44%	0.989	58%
$H$	0.92		0.93		0.96		0.95		0.87	
EI										
$\sigma_G^2$	1.368**	16%	1.018*	16%	0.683*	14%	0.538*	11%	0.835**	16%
$\sigma_{GL}^2$	1.482***	17%	2.727***	42%	1.130***	23%	1.156***	24%	1.224***	23%
$\sigma_{GC}^2$					0.347**	7%	0.088*	2%	0.120*	2%
$\sigma_{GLC}^2$					0.118 <sup>NS</sup>	2%	0.392***	8%	0.306***	6%
$\sigma_e^2$	5.62	66%	2.792	43%	2.730	55%	2.671	55%	2.886	54%
$H$	0.77		0.68		0.64		0.69		0.70	

NS nonsignificant

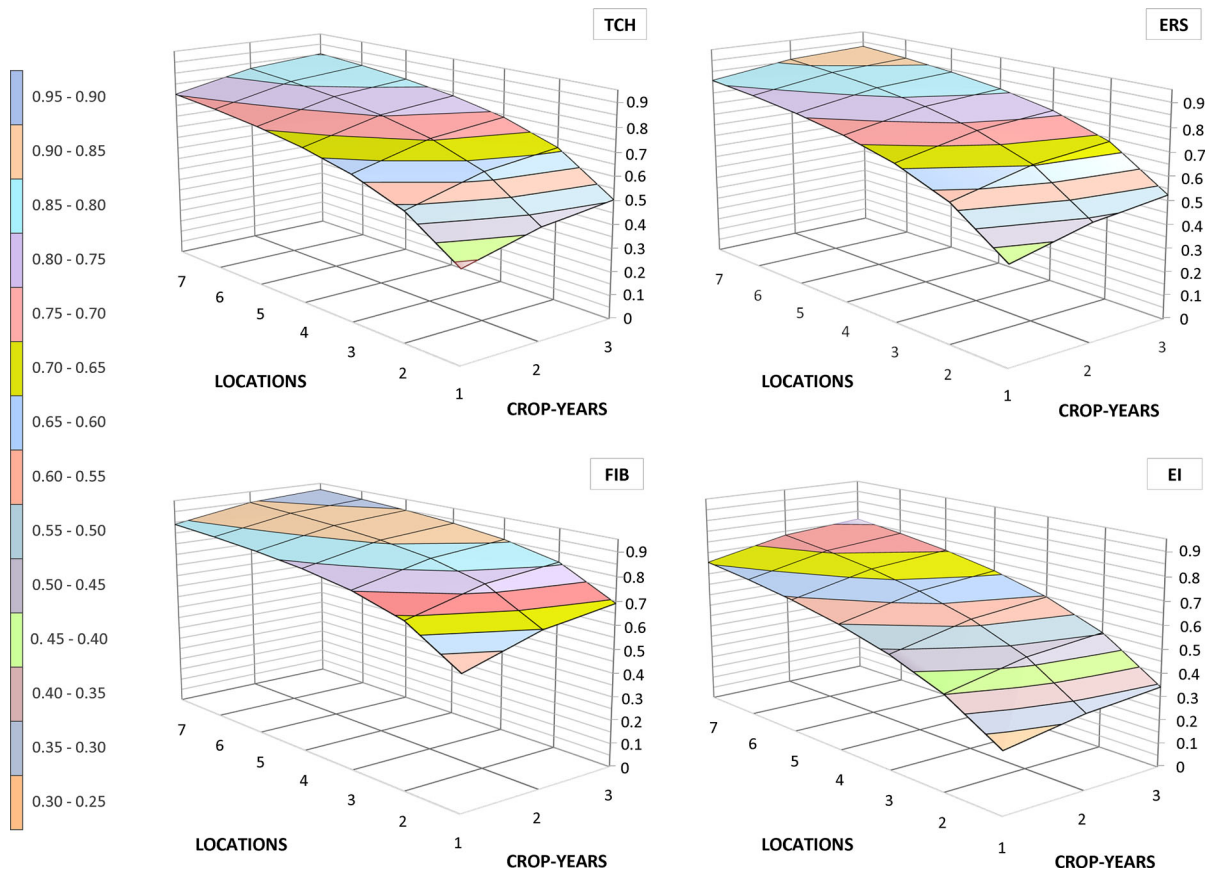
\* ( $\text{Pr} > Z$ )  $\leq 0.05$ \*\* ( $\text{Pr} > Z$ )  $\leq 0.01$ \*\*\* ( $\text{Pr} > Z$ )  $\leq 0.001$ 

less predictable trait (EI). Location generated a slightly greater increase of  $H$  compared with crop year: depending on the trait considered, an increase of two units in location ( $L = 3$  compared with  $L = 1$ ) resulted in an increase in the range of 0.20–0.26 in  $H$  but only 0.08–0.14 when considering a similar increase in crop years ( $C = 3$  compared with  $C = 1$ ).

#### Statistical analysis of GGE data

The adjusted means of the 21 genotypes of the two oldest series (S00 and S03) which were tested in common at the seven locations across two crop years (Tables 5, 6) represented the balanced GGE data further used to construct biplots representative of





**Fig. 1** Simulated broad-sense heritability ( $H$ ) of tonnes of cane per hectare (TCH), estimable recoverable sugar (ERS), fiber content (FIB), and economic index (EI) expressed as a function

of their variance components (model 2) and combinations of various numbers of locations and crop years

genotype  $\times$  environment (GE) relations. This genotype set, which gathered elite candidates previously selected from one of the seven contrasted locations, logically exhibited (i) a relatively high 20.6% coefficient of variation (CV) mean between genotypes for TCH, (ii) more moderate CV means for ERS (7.8%) and FIB (8.9%), and (iii) a subsequent relatively high CV of 20.9% for the derived EI trait. The adjusted means of the 21 genotypes for all locations were subjected to analyses of variance and the trait means for the locations were compared using a Duncan's multiple-range test at  $P = 0.05$ . ANOVA revealed a highly significant ( $P < 0.0001$ ) variance for location for each of the four yield-related traits (TCH, ERS, FIB, and EI). Statistical ranking of location means varied a lot between TCH ( $SB > ES \geq LM \geq GL > MN = SP = VB$ ), ERS ( $LM > GL = SP = VB > MN = SB > ES$ ), FIB

( $GL > MN \geq LM \geq ES = VB > SB > SP$ ), and EI ( $LM \geq SB \geq GL > ES = VB > MN = SP$ ) and did not reveal any broad similarity between any two traits. Correlations of traits across locations revealed in the set of genotypes surveyed are presented in Table 7. There was no correlation between TCH and FIB or between ERS and EI, high positive correlation between TCH and EI (0.746), high negative correlation between TCH and ERS ( $-0.626$ ), and moderate negative correlations between ERS and FIB ( $-0.367$ ) and between FIB and EI ( $-0.342$ ). The ecovalence of the seven locations ( $W_j$ ) varied around its average value ( $14\% = 100\%/7$  locations) with larger range (rg) for TCH (rg = 20.9%) and ERS (rg = 24.9%) compared with FIB (rg = 12.9%) or EI (rg = 13.4%). For TCH, the smallest contribution to genotype  $\times$  location interaction was due to location LM ( $W_{LM} = 5.3\%$ ) and the highest to location SB

**Table 5** Adjusted mean for tonnes of cane per hectare (TCH) and estimable recoverable sugar (ERS) over the two first crop years of the 21 genotypes of series S00 and S03 tested at the seven locations along with location or genotype ecovalence indexes and variation between genotypes

Genotype	TCH (Mg ha <sup>-1</sup> )		ERS (%)																		
	Location		Mean		W <sub>m</sub> (%) <sup>1</sup>		Location										Mean		W <sub>m</sub> (%) <sup>1</sup>		
	ES	GL	LM	MN	SB	SP	VB	ES	GL	LM	MN	SB	SP	VB	ES	GL	LM	MN	SB	SP	VB
G1	99.8	106.8	118.4	49.0	136.8	60.4	88.8	94.3	3.7	9.7	11.3	12.2	8.7	11.0	11.1	11.7	10.8	7.3			
G2	91.2	80.7	102.8	72.7	122.4	75.5	107.5	93.2	4.9	10.1	12.2	13.5	10.9	12.9	12.9	13.0	12.2	6.7			
G3	110.2	94.0	88.6	36.0	108.2	69.8	50.4	79.6	3.5	11.0	11.6	12.4	11.1	12.0	12.2	12.2	11.8	1.8			
G4	122.9	111.3	124.7	79.1	157.8	70.7	116.6	111.9	3.0	10.6	11.4	12.3	10.3	11.1	11.9	12.1	11.4	1.6			
G5	101.7	99.8	92.5	74.2	144.4	56.1	55.0	89.1	3.0	10.8	10.9	12.1	10.8	11.0	11.5	11.7	11.3	1.8			
G6	112.4	75.0	125.6	92.8	148.6	34.5	92.8	97.4	9.9	9.8	11.6	11.5	10.4	11.7	12.2	12.2	11.4	3.9			
G7	96.2	97.1	107.4	39.6	149.0	54.5	65.2	87.0	4.9	10.2	11.5	11.9	7.8	10.5	11.1	11.7	10.7	14.5			
G8	106.4	100.6	123.6	73.8	139.0	78.3	84.1	100.8	1.1	9.8	11.9	12.5	10.7	10.9	12.1	12.7	11.5	2.5			
G9	125.5	116.7	144.0	128.4	195.8	73.4	135.6	131.3	11.2	9.1	9.5	9.0	10.3	8.8	11.8	11.2	10.0	15.2			
G10	138.8	114.8	132.6	96.2	180.3	104.2	136.5	129.1	3.7	9.7	10.7	10.7	11.7	9.7	11.6	11.9	10.8	7.5			
G11	113.6	113.7	89.2	38.8	62.6	82.8	63.0	80.5	15.7	10.2	12.6	13.1	11.2	10.9	11.3	12.0	11.6	3.2			
G12	131.1	103.8	109.3	84.2	148.4	92.9	87.8	108.2	0.7	8.9	10.5	11.3	10.9	9.8	10.4	10.4	10.3	2.4			
G13	127.4	91.2	104.8	101.1	159.8	94.0	87.2	109.4	3.4	9.4	10.6	10.8	10.3	10.2	10.8	10.8	10.4	1.0			
G14	123.9	109.1	100.6	48.7	91.6	73.1	64.2	87.3	6.6	10.8	12.6	13.0	11.9	11.7	12.3	10.9	11.9	4.6			
G15	138.2	104.2	108.6	84.9	111.8	81.1	75.8	100.6	3.6	9.5	10.9	12.0	11.1	9.6	12.3	10.9	10.9	4.2			
G16	138.0	118.7	111.6	91.6	148.2	97.0	70.0	110.7	2.8	10.1	11.9	12.8	11.0	10.6	10.8	11.2	11.2	3.2			
G17	119.5	97.4	93.4	105.1	129.5	76.2	65.2	98.0	4.5	8.9	10.8	12.2	10.3	10.0	10.8	10.8	10.6	1.6			
G18	143.6	116.5	113.1	101.2	159.8	89.0	85.0	115.4	1.9	9.5	11.2	12.6	11.7	9.3	12.1	10.9	11.0	7.1			
G19	156.9	121.8	125.8	103.1	128.7	124.3	93.8	122.1	6.2	9.4	11.4	12.2	10.3	9.8	10.4	9.9	10.5	4.5			
G20	126.7	109.8	119.5	83.4	136.1	98.4	79.6	107.6	1.3	9.4	10.5	12.0	10.7	9.5	10.7	9.7	10.4	4.3			
G21	111.8	119.9	116.1	81.0	166.3	85.9	123.8	115.0	4.5	8.9	9.9	10.8	10.1	9.8	10.7	10.3	10.1	0.9			
Mean <sup>2</sup>	121.87b	104.9c	112.0bc	79.3d	139.3a	79.6d	87.0d	103.4		9.8d	11.2b	11.9a	10.6c	10.5c	11.5b	11.3b	11.0				
CV(g) (%) <sup>3</sup>	13.7	12.0	13.1	31.3	21.3	24.6	28.6	20.6		6.5	7.2	8.4	8.8	9.7	6.4	7.8	7.8				
W <sub>j</sub> (%) <sup>4</sup>	10.9	11.1	5.3	13.8	26.2	16.6	16.1			6.4	6.6	17.6	31.3	13.0	10.4	14.6					

<sup>1</sup> Genotype ecovalence stability index<sup>2</sup> Means with the same letter are not significantly different ( $P > 0.05$ )<sup>3</sup> Coefficient of variation between genotypes<sup>4</sup> Location ecovalence stability index

**Table 6** Adjusted mean for fiber content (FIB) and economic index (EI) over the two first crop years of the 21 genotypes of series S00 and S03 tested at the seven locations along with location or genotype ecovalence indexes and variation between genotypes

Genotype	FIB (%)		EI															
	Location							Mean							W <sub>m</sub> (%) <sup>1</sup>			
	Location							Mean							W <sub>m</sub> (%) <sup>1</sup>			
	ES	GL	LM	MN	SB	SP	VB	ES	GL	LM	MN	SB	SP	VB	Mean	W <sub>m</sub> (%) <sup>1</sup>		
G1	14.9	16.6	15.6	15.7	14.1	14.3	14.5	15.1	3.0	6.7	9.7	11.2	3.4	10.5	5.6	8.1	7.9	3.9
G2	13.9	15.6	14.9	15.1	13.1	13.7	13.4	14.2	4.0	6.4	8.2	11.5	5.6	11.2	9.3	11.4	9.1	7.4
G3	15.5	16.7	15.9	16.3	14.3	14.6	15.8	15.6	1.1	8.6	8.9	8.4	3.1	8.6	5.7	4.9	6.9	3.5
G4	13.3	14.2	13.8	13.5	12.6	13.0	13.2	13.4	1.6	9.1	9.8	11.8	6.2	12.2	7.2	11	9.6	2.7
G5	15.0	15.5	14.3	15.0	12.8	13.9	14.2	14.4	1.9	7.9	9.1	8.9	5.3	11	6.7	5.1	7.7	3.4
G6	12.4	13.7	12.0	13.3	12.3	12.8	12.9	12.8	5.8	7.2	6.9	10.7	6.3	11.9	4.7	8.9	8.1	5.7
G7	13.2	14.8	12.8	16.1	11.9	13.8	14.4	13.9	17.7	6.9	8.9	9.6	2.1	10.5	5.5	6	7.1	4.6
G8	13.6	14.1	14.0	13.9	12.6	13.3	13.2	13.5	1.8	7.1	9.8	11.9	5.2	10.1	6.4	8.4	8.4	1.8
G9	13.9	13.9	14.5	15.4	12.7	12.8	14.4	14.0	7.5	7.3	8	8.6	8.4	11.4	7	11.5	8.9	8.4
G10	14.6	15.3	14.7	14.4	13.4	13.4	14.2	14.3	1.9	9.1	9.4	10	7.5	11.7	9.1	12.5	9.9	5.6
G11	14.2	14.9	14.4	14.4	12.8	13.5	13.3	13.9	1.7	8.1	11.6	9.2	2.6	4.8	5.1	6.3	6.8	12.0
G12	17.8	18.3	17.4	17.2	15.7	16.8	16.3	17.1	4.6	7.1	7.5	9.3	6.4	9.2	5.7	7.1	7.5	0.9
G13	16.2	17.8	16.3	16.2	15.2	15.8	17.2	16.4	4.1	7.7	6.9	8.5	7.4	10.7	6.5	7.4	7.9	3.7
G14	13.8	15.5	14.0	13.8	12.4	14.5	14.2	14.0	5.7	10	11.2	10.5	4	7.9	5.7	5.8	7.9	7.1
G15	14.4	15.0	14.6	14.5	12.7	14.7	14.5	14.3	3.1	8.7	8.4	10.1	7.1	6.4	7.3	6.8	7.8	4.7
G16	15.6	15.9	15.1	15.0	14.0	14.9	14.9	15.0	2.2	9.8	11	11.7	7.5	10.7	6.3	6.4	9.1	3.4
G17	15.9	17.2	17.9	16.5	15.2	15.8	17.5	16.6	8.1	6.6	7.5	8.8	7.8	8.2	7.6	5.8	7.5	4.5
G18	14.2	15.5	14.5	14.4	12.5	14.0	15.6	14.4	6.0	9.2	10	11.6	9.1	9	8.3	7.3	9.2	3.6
G19	14.4	15.7	15.3	14.7	12.9	12.2	13.7	14.1	10.3	9.5	10.8	12.2	7.7	7.9	10.6	7	9.4	7.3
G20	17.1	17.6	16.2	17.6	15.6	17.1	17.2	16.9	4.5	7.9	8.2	11.2	6.4	7.9	6.8	6	7.8	2.4
G21	13.8	14.2	14.4	14.9	12.1	13.1	14.0	13.8	3.2	6.4	8.8	9	5.9	11	5.8	9.3	8.0	3.2
Mean <sup>2</sup>	14.7c	15.6a	14.9bc	15.1b	13.4e	14.2d	14.7c	14.7		8c	9.1b	10.2a	6d	9.7ab	6.8d	7.8c	8.2	
CV(g) (%) <sup>3</sup>	9.0	8.5	9.2	7.7	8.9	9.2	9.5	8.9		14.6	15.1	12.7	32.9	20.2	22.0	28.8	20.9	
W <sub>j</sub> (%) <sup>4</sup>	7.9	9.5	18.6	20.8	8.1	16.9	18.3			9.7	15.6	7.7	17.2	20.4	8.2	21.1		

<sup>1</sup> Genotype ecovalence stability index<sup>2</sup> Means with the same letter are not significantly different ( $P > 0.05$ )<sup>3</sup> Coefficient of variation between genotypes<sup>4</sup> Location ecovalence stability index

**Table 7** Global phenotypic correlations between the four yield-related traits computed on the mean value across locations of the balanced genotype dataset used for the GGE study

Trait	TCH	ERS	FIB	EI
TCH	1			
ERS	−0.626	1		
FIB	−0.006	−0.367	1	
EI	0.746	−0.002	−0.342	1

( $W_{SB} = 26.2$ ). For ERS, the smallest contribution to the interaction was due to ES ( $W_{ES} = 6.4\%$ ) and GL ( $W_{GL} = 6.6\%$ ) and the highest to MC ( $W_{MC} = 31.3\%$ ). Similarly, for each trait, the individual contribution of genotypes to the genotype  $\times$  location interaction was variable around the average ecovalence value of  $W_m = 4.76\%$  (100%/21 genotypes). However, the individual contribution of genotype remained rather modest, since in 85% (77/84) of all cases (four traits considered together)  $W_m$  did not exceed 7% and in the remaining 15% of cases the highest genotype ecovalence reached a maximum of 17.7% (genotype G7 for FIB).

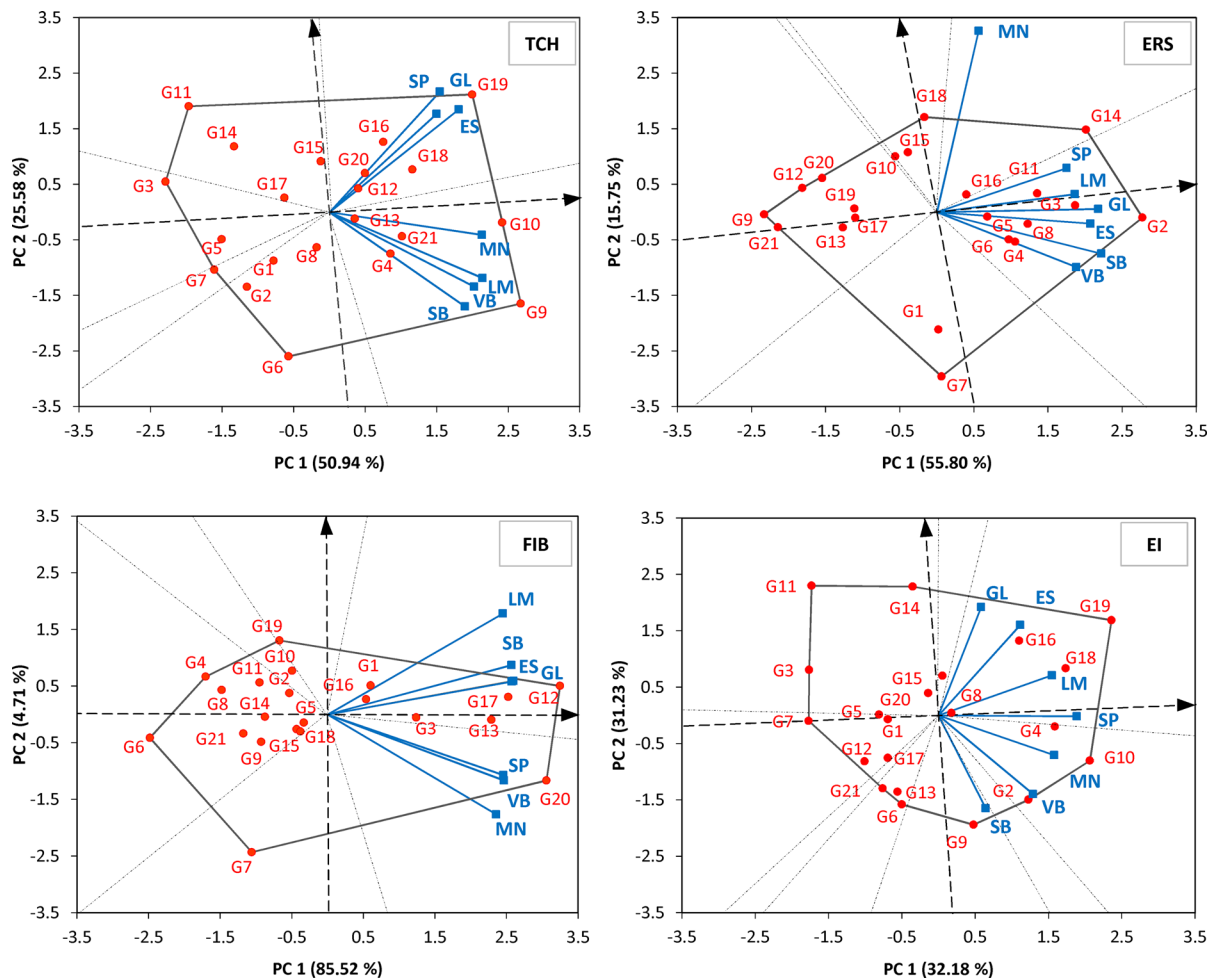
#### Overview of the distribution of genotypes in GGE biplots

The symmetric scaling of genotype and environment scores in GGE biplots permits direct visualization of the magnitude of genotypic and environmental variations in the same units for both PC1 and PC2 (Yan 2002). Figure 2 presents GGE biplots of the four traits. As a general rule, for all traits, environment vectors always had positive abscissa and the variation among them was first discriminated by PC2 while variation among genotypes was first discriminated by PC1. These first two principal components (PCs) of biplots explained 76.52, 71.55, 90.23, and 63.41% of the total GGE variation of TCH, ERS, FIB, and EI, respectively (Fig. 2). This suggests that a biplot represented by both PC1 and PC2: (i) adequately approximates the GGE data of TCH and ERS, (ii) represents very accurately the GGE data of FIB, but (iii) represents less efficiently the GGE data of EI. For this latter trait, data variability can be represented with a level roughly similar to the other three biplots when taking into account the additional contribution of its PC3 (15.16%). The polygon formed by connecting the

genotypes that are further away from the biplot origin contains all genotypes. The wider spread of genotypes over the whole biplot plan for both TCH and EI as opposed to both ERS and FIB reflects the much higher contribution of the respective PC2 for the two former characters (25.58 and 31.23%) compared with the latter two (15.75 and 4.71%) in the variation of their respective GGE data. Comparing the distribution pattern of genotypes in the biplots of TCH and ERS showed abscissa of opposite sign for a very large number of genotypes (17/21) between the two traits. The high frequency of this opposite direction of genotype abscissa between TCH and ERS is in agreement with (i) the negative correlation coefficient between these two traits (−0.626), and (ii) the similarity of the weight of their PC1 axes (50.94 or 55.80%) in the variation of GGE data. Finally, the coordinates of genotypes in TCH and EI biplots were of identical sign for a majority of 14 of the 21 abscissa (8 negative and 6 positive) and a majority of 16 of the 21 ordinates (8 positive and 8 negative). These double findings are perfectly congruent with (i) the positive correlation observed between these two traits (0.746) and (ii) the similar order of magnitude of the contribution of both their PC1 (50.94 and 32.18%) and PC2 (25.58 and 31.23%) axes in the variation of their GGE data.

#### The which-won-where pattern

The format of the polygons surrounding all genotypes displays the which-won-where pattern and hence is a succinct summary of the GE pattern of the MET dataset. The rays that are perpendicular to the sides of the polygon (or their extensions) divide the TCH, ERS, FIB, and EI biplots into six, six, six, and nine sectors, respectively, allowing immediate visualization of the similarity between locations for the genotype response. An interesting feature of this view of a GGE biplot is that the genotype on the vertex for each sector had the highest trait value in all environments that fall in the sector, providing that the percentage of GGE data explained by the biplot is high enough to accurately reflect the original data (Yan et al. 2000; Yan 2002). For instance, for the TCH biplot that explained 76.52% of the GGE data, the seven environments fell in one sector for three of them (SP, GL, and ES). The four remaining environments (MN, LM, VB, and SB) fell in the second sector. The



**Fig. 2** GGE biplots of the four yield-related traits (TCH, ERS, FIB, and EI) based on data of 21 genotypes of series S00 and S03 tested at seven locations. PC1 and PC2 are the principal component scores on the first and second axis, respectively. The variation accounted for by the axes is shown in *brackets*. Biplots are based on symmetric scaling between genotypes and

environments (scores with the same units for both PC1 and PC2). Environments are indicated by two letters, and genotypes (G) by numbers. The reference frame indicated by *dashed arrows* represents an average environment vector and its orthogonal vector

vertex genotypes for these two sectors were G19 and G9, respectively, suggesting that these two genotypes had the highest tonnage or nearly the highest in all environments that fell in each respective sector. In fact, both G19 and G9 were by far the best yielding genotypes in all locations falling in their respective sectors (Table 5), except in VB where G9 was the second best genotype just after G10 (with only  $1.1 \text{ Mg ha}^{-1}$  difference). The reliability of the which-won-where pattern revealed in biplots by vertex genotypes depends on (i) the cumulated weights of PC1 and PC2 axes, and (ii) the narrowness of the angles that encompass all the vectors of environments

falling in each sector of interest. For example, in the biplot of FIB, the weight of the first two axes was rather high (90.23%) and both angles in the two sectors that grouped either four or three locations were each relatively acute. Congruently, the which-won-where pattern identified the genuine highest variety in five environments (G12 in ES, GL, and SB; G20 in MN and SP) but the second highest one in two environments (G12 in LM; G20 in VB). In the biplot of ERS, the first two axes cumulated a weight (71.55%), slightly smaller than that of the FIB biplot, and the angle that encompassed almost all environment vectors (6/7) in a single sector was slightly wider than the

two angles in the FIB biplot. Congruently, the which-won-where pattern identified the genuine highest variety in five environments (G2 in LM, SB, SP, and VB; G14 in MN) but the third best in one environment (G2 in GL), and the seventh best in the remaining environment (G2 in ES). In the biplot of EI, the total weight of both first axes dropped to 63.41% of the total GGE variation and the angle that encompassed four of the seven environment vectors in the same sector (GL, ES, LM, and SP) was significantly wider than the angles examined in all of the other three biplots. Congruently, the which-won-where pattern identified the genuine best variety only in three environments (G19 in LM and SP; G10 in VB) but the second best in one environment (G19 in ES), the third best in one environment (19 in GL), and the fourth best in the two remaining environments (G10 in MN; G9 in SB). To obtain a more global view of the ability of biplots to display reliable classifications of the performance of all genotypes, we calculated a correlation coefficient between the genotype ranking along each environment axis and the corresponding genotype ranking in the initial data (Tables 5, 6) and computed an overall mean across all seven environments. This mean correlation coefficient for genotype ranking reached 0.93 for FIB, 0.84 for both TCH and ERS, and 0.50 for EI. The hierarchy of these four coefficients of correlation (FIB > TCH = ERS > EI) and the magnitude of their relative differences are perfectly in alignment with the hierarchy and differences observed in the percentage of variation, respectively, explained by the biplots of the four traits (90.23% > 76.52%  $\approx$  71.55% > 63.41%).

#### Interrelationship among environments

Correlation coefficients among the seven environments are presented in Tables 8 and 9 with bold characters indicating values that are statistically different from zero ( $P < 0.05$ ). The vector view of GGE biplots provides a succinct summary of the interrelationships among environments. For instance, in the biplot of TCH, the seven environments fell in two sectors that grouped either three or four environments. More or less acute angles between vectors could only be observed between environments within a same sector, congruently with the facts that (i) all correlation coefficients between any two locations within each sector were significant, and (ii) conversely

**Table 8** Phenotypic correlations of TCH (below diagonal) and ERS (above diagonal) between locations inferred from the balanced dataset used in the GGE study

Location	ES	GL	LM	MN	SB	SP	VB
ES	X	<b>0.66</b>	<b>0.52</b>	0.13	<b>0.71</b>	<b>0.52</b>	<b>0.55</b>
GL	<b>0.58</b>	X	<b>0.85</b>	0.18	<b>0.73</b>	0.40	<b>0.51</b>
LM	0.34	0.31	X	0.19	<b>0.62</b>	0.26	0.28
MN	<b>0.58</b>	0.19	<b>0.57</b>	X	0.03	0.35	−0.03
SB	0.17	0.14	<b>0.70</b>	<b>0.66</b>	X	<b>0.53</b>	<b>0.70</b>
SP	<b>0.74</b>	<b>0.62</b>	0.17	0.37	0.05	X	<b>0.67</b>
VB	0.18	0.25	<b>0.80</b>	<b>0.54</b>	<b>0.67</b>	0.22	X

Bold figures indicate correlations significantly different from zero at  $P < 0.05$

**Table 9** Phenotypic correlations of FIB (below diagonal) and EI (above diagonal) between locations inferred from the balanced dataset used in the GGE study

Location	ES	GL	LM	MN	SB	SP	VB
ES	X	<b>0.62</b>	0.34	0.21	− <b>0.24</b>	0.28	−0.16
GL	<b>0.92</b>	X	<b>0.44</b>	−0.27	− <b>0.34</b>	0.11	−0.16
LM	<b>0.88</b>	<b>0.87</b>	X	0.17	<b>0.08</b>	0.39	0.20
MN	<b>0.81</b>	<b>0.79</b>	<b>0.73</b>	X	<b>0.19</b>	<b>0.58</b>	0.33
SB	<b>0.93</b>	<b>0.91</b>	<b>0.88</b>	<b>0.78</b>	X	0.03	<b>0.60</b>
SP	<b>0.87</b>	<b>0.85</b>	<b>0.72</b>	<b>0.76</b>	<b>0.85</b>	X	0.35
VB	<b>0.82</b>	<b>0.83</b>	<b>0.79</b>	<b>0.80</b>	<b>0.82</b>	<b>0.85</b>	X

Bold figures indicate correlations significantly different from zero at  $P < 0.05$

correlations between any two environments belonging to different sectors were not significant, with the sole exception of the correlation between MN and ES [which can easily be explained by the coordinates of these two environments on PC3 with significant value and identical positive sign (data not shown)]. For the ERS biplot, there is a major separation in two sectors between MN on one side and all other six environments in the other side, with a mean angle vector roughly about a right angle. This biplot view accurately reflects the absence of any significant correlation between MN and the other six locations (Table 8). Regarding the two-by-two relationships among these other six environments as visualized by couples of vectors showing more or less acute angles, they all logically exhibit a significant correlation except three particular couples (LM–SP, GL–SP, and LM–VB). These particular couples of environments had



coordinates of opposite sign (negative for SP and VB, positive for GL and LM) on the PC3 axis (data not shown), and the fact that the weight of this axis (14.92%) was marginally less than that of PC2 (15.75%) logically explains the genuine angle values higher than in the two-dimension biplot and therefore loose associations. In the FIB biplot, the percentage of the total GGE variation explained by the first component axis was high (85.52%) as well as the abscissa value of all environment vectors. The combination of these two features suggests coefficients of correlation that are not only significant but also probably high and positive in all cases. This picture was perfectly supported by the correlation values ranging between 0.72 and 0.93 (Table 9). Finally, the EI biplot revealed a large fan-shaped distribution of environment vectors from negative (SB) to positive (LG) ordinates onto a PC2 axis that showed a significant weight (31.23%) hardly inferior to that of PC1 (32.18%). This picture suggests a majority of modest correlation coefficients of either positive or negative sign with only a few, likely positive significant correlations. The values perfectly supported these suggestions (Table 9). Only four positive correlation values were indeed significant (GL–ES, SP–MN, VB–SB, and GL–LM). They logically corresponded for three of them to couples of environments represented by vectors displayed in immediate neighbor positions.

### Mean yield and stability of genotypes

Visualization of both mean performance and stability of genotypes across all environments is always an important issue in cultivar evaluation and to assist in decision-making for selection purposes. This can be done by following the methodology proposed by Yan (2002), which consists in the introduction of an average environment (AE) vector in trait biplots whose coordinates are defined by the average of PC1 and PC2 scores of all environments. This AE vector and its orthogonal vector define a new orthogonal reference frame in which: (i) the highest positive (or negative) abscissa in the new frame should pinpoint the genotypes exhibiting the highest (or lowest) mean performance across all environments, and (ii) the highest ordinates in absolute value (either positive or negative) in this new frame should pinpoint the genotypes exhibiting the highest instability of their

performance across environments. These guidelines for analysis in this new reference frame (pictured by dashed arrows in Fig. 2) permits one to identify at a glance: (i) three (G9, G10, G19), one (G2), two (G12, G20), and two (G10, G19) genotypes likely having the highest overall performance means for TCH, ERS, FIB, and EI, respectively, and (ii) one (G3), one (G9), one (G6), and three (G3, G7, G11) genotypes likely having the lowest overall performance means for the same respective traits. All these graphical findings provided by the four two-dimension biplots are perfectly supported by the genotype mean values observed in the original data (Tables 5, 6). They support the reliability of the AE reference frame in displaying a representative ranking of best and poorest mean performances, although biplots did not explain 100% of GGE variation. Moreover, genotypes G11, G7, G7, and G11 appeared to have the highest ordinates in absolute values in this new reference frame for TCH, ERS, FIB, and EI, respectively. These graphical findings pinpoint genotypes that should exhibit the poorest or almost the poorest performance stability across environments. Comparison of these findings with ecovalence values of genotypes ( $W_m$ ) that measure the contribution of each genotype to the GE interaction (Tables 5, 6) were very congruent. For three of the four traits (TCH, FIB, and EI), the graphical view indeed identified the most unstable genotype, and for the remaining one (ERS) the second most unstable genotype.

### Discussion

The present study dissected genotype  $\times$  environment interactions of traits related to yield components of sugarcane that might exist in the context of sugarcane cultivation on Réunion Island. To this end, we analyzed data from four MET series tested recently in the final stage of eRcane's selection program based on a network of seven locations representative of the main cultivation zones of sugarcane on Réunion. Genotype-by-environment data of four yield-related traits were investigated to study the main characteristics of genotype response in the multilocation selection scheme. Data were analyzed using mixed linear models to estimate variance components and explored in detail using GGE biplots displaying information on both genotypes and environments.

### Importance of the different variance components

Results obtained from our second, “global” model provide estimates of variance components on the basis of the analysis of material belonging to four recent selection series representative of the current breeding program of eRcane. This reference analysis revealed a genetic variance ( $G$ ) higher than each variance component of the interactions ( $GL$ ,  $GC$ , and  $GLC$ ) for all traits, except for the key  $EI$  trait, for which the  $GL$  component was substantially greater than  $G$ . Concretely, the  $GL$  interaction reflects the mean magnitude of changes in genotype rankings across locations; the  $GC$  interaction represents the importance of the fluctuations of performance of genotypes across crop years, which is related to their ratooning ability; the  $GLC$  interaction reflects the ratooning ability of genotypes as influenced by location. For all traits, the  $GL$  interaction was more important than  $GC$ , indicating that testing genotypes across locations is more important than testing for ratooning ability. Besides,  $TCH$  and  $EI$  produced a  $GLC$  highly significantly ( $P < 0.001$ ) larger than  $GC$ . This indicates that the ratooning ability of the genotypes is location specific and illustrates the relative complexity of selecting for these traits, which must be assessed at each location. On the contrary, the null value of the  $GLC$  component for  $FIB$  indicates that the evolution of  $FIB$  across crop years tends to be similar from one location to another.

Simulation of broad-sense heritability ( $H$ ) with increasing numbers of locations ( $L$ ) and crop years ( $C$ ) revealed that: (i) accurate assessment of the mean value of  $FIB$  across all environments can be obtained with relatively modest experimental effort ( $L + C \leq 3$ ), (ii) the mean performance of genotypes across environments for  $TCH$  and  $ERS$  required greater experimental effort ( $L + C > 4$ ) than  $FIB$  to reach similar accuracy levels, and (iii) for  $EI$ , the lower weight of its genetic variance ( $G$ ) compared with its  $GL$  variance determined  $H$  values lower than for the other three traits. For all the traits, the reliability of estimates of genotype means ( $H$  values) was influenced more markedly by an increase in the number of locations than by an increase in the number of crop years.

Among the four traits,  $EI$  is the most important selection criterion for growers and millers since it represents an economic profitability index. Therefore,

to rapidly grasp and compare the potential merit of different candidate genotypes in the diverse cultivation context of Réunion Island, it appears more important to place effort first on the number of trial locations, by identifying additional locations likely to increase the representativeness of the basic location network. In case of budget constraints on resources allocated to field experiments, it would be much more advisable to reduce the number of crop cycles to favor the number of test locations.

### Efficiency of GGE plots for visualizing MET data

To evaluate the sugarcane MET network of Réunion Island and obtain essential information for some of the most promising candidate genotypes, we tested the methodology of GGE biplot tools (Yan and Tinker 2006) on a balanced dataset from the two most advanced series ( $S00$  and  $S03$ ). The two first components of standardized GGE biplots explained a particularly high percentage of initial GGE data for  $FIB$  (90%), relatively high for  $TCH$  (77%) and  $EI$  (72%), and moderately high for  $EI$  (63%). Two-dimensional GGE biplots of genotypes and locations sufficiently approximated initial data, providing: (i) a convincing congruency between the comparative arrangement of genotypes on biplot planes of two traits linked by significant positive ( $TCH$  and  $EI$ ) or negative ( $TCH$  and  $ERS$ ) correlations, (ii) easy and reliable visualization of the best performing candidates in each environment, (iii) good immediate visualization of the ranking of genotypes in each environment with an accuracy level directly related to the percentage of initial variation explained for each trait, (iv) a succinct summary of interrelationships among environments, in which acute angles between environment vectors in the same sector depicted at least correlations between environments that are significant when GGE variation is approximately represented by biplots ( $TCH$ ,  $ERS$ ,  $EI$ ), or actually high correlations when variation is represented very efficiently ( $FIB$ ), and (v) a rapid and broad-brush view of the ranking of mean performance and stability of genotypes across all environments within a new orthogonal reference frame based on a vector of a virtual average environment. All these graphical findings inferred from GGE biplots illustrate the value of these tools to interpret and visualize data results at a glance.

## Genotypic response across the multilocation selection network on Réunion Island

Analyses of different genotype-by-location datasets using both mixed models and GGE biplots revealed the general characteristics of genotypic response within the multilocation selection network of Réunion Island and provided valuable information about the selection scheme. Variability between candidate genotypes at this multilocation stage is relatively important for TCH (CV = 20.1%) considering their diverse selection origins from contrasted agroclimatic environments and the original large variation of the trait in earlier selection stages (Dumont pers. commun.). On the contrary, the variability appeared more modest for the ERS (CV = 7.8%) and FIB (CV = 9.8%) quality traits because of the more modest genetic variations than for TCH in the eRcane breeding germplasm and early selection pressure put on ERS since the beginning of selection (Hoarau, pers. commun.). For EI, its variation in this final selection stage appeared relatively high (CV = 20.9%) due to its derived computation from ERS and TCH. Genotype response differed markedly for FIB on one side, for TCH and ERS on a second side, and for EI on a third side. For FIB, the very limited weight of all interaction components (GL, GC, and GLC) compared with the genetic component (G) is reflected by considerably high broad-sense heritability. This implies a high positive correlation between all locations (from 0.72 to 0.92), and many environments are perfectly redundant for genotype ranking for this trait. However, for TCH and ERS, significant coefficients of correlation between environments are much less frequent (10 or 12 occurrences among 21, respectively). Moreover, correlations reached much lower value ranges than for FIB, in alignment with lower heritability levels. Therefore, the chance and strength of redundancy between environments for genotype ranking for each of these two traits are further reduced. Indeed, simultaneous significant correlated couples of environments for both traits (ES/GL, ES/SP, LM/SB, SB/VB) exhibited correlation coefficients in a range not high enough (0.52–0.70) to provide identical or similar genotype ranking among couples of environments for both traits, particularly in the top elites.

## Selection strategy for the economic index

EI is the most important criterion for selection for both sugarcane producers and industry, since it represents an economic profitability index. Compared with the other traits, EI is more poorly heritable at the whole MET level (0.70) because of interaction variance components of higher importance. The large part of GGE variation left unexplained by its GGE biplot (more than one-third) combined with the large fan-shaped distribution of environment vectors reflects a nonsignificant or generally loose association between any two environments, as supported by the correlation coefficients (Table 9). This situation indicates an absence of any remarkable similarity of genotype rankings between environments and therefore underlines the importance of the whole selection network. In this context, identification of genotypes adapted to specific environments rather than genotypes with broad adaptation appears to be the most efficient selection strategy to pursue to achieve global genetic progress at the level of the whole sugarcane industry. However, three couples of environments, namely ES/GL, MN/SP, and SB/VB, exhibited associations that were not complete loose, since their correlations for the EI trait ranged from 0.58 to 0.62. Correlations between ES and GL can be explained by the fact that these two sites are separated by a short distance and share the same soil type, differing only in their irrigation regime (Table 1). Correlation between MN and SP can be explained by these sites being relatively unfavorable for cane cultivation with similar chemical fertility of their soil (low pH, low cation exchange capacity) as well as their physical characteristics (very stony), which could imply some similarities in the selection pressure exerted on the tested material. Finally, correlation between SB and VB cannot be obviously explained, since these two environments did not share any similar agroclimatic parameters.

These three particular correlations represent useful information likely to help in rationalization of varieties to be tested in this final MET selection stage. Each year, 5–10 varieties are routinely found to be superior to local controls at each of the seven stations in the previous selection stage. All these elite varieties cannot be tested everywhere due to constraints on land resources and budget. The choice of material entering the MET testing stage always implies difficult

decisions between multiple potentially elite genotypes. These modest but interesting correlations between ES and GL, MN and SP, and SB and VB suggest that better genetic gains per unit cost (in the subsequent variety release) could be obtained if greater attention is paid to the choice of a superior proportion of common genotypes between these three couples of environments.

## Conclusions

Selection of elite sugarcane genotypes for cultivation on Réunion Island relies on a MET selection program conducted by eRcane research institute. This MET program has been progressively developed over two decades at an increasing number of locations that are representative of the different agroclimatic zones used for sugarcane production. Nowadays, this work routinely includes an unprecedented number of seven experimental stations. Since 2011, each annual new elite series has been systematically tested at all stations of this basic network using a robust standard trial design (in terms of plot size and number of replications) and during an adequate number of crop years to test ratooning ability. Depending on opportunities for collaboration with some producers, some sugarcane series may occasionally be partially tested in a few additional trials at varying locations.

This work is the first report on the MET selection program on Réunion Island in its present-day dimension. We analyzed data so far available from the first four series been tested across the full seven-station network. The choice of the location of these seven stations was guided by the search for the most contrasting areas as possible on the agroclimatic map of the local sugarcane industry, in the hope of efficiently selecting for local adaptation. Data analyses revealed highly significant genotype  $\times$  location (GL) interaction for all traits of economic interest, as well as for genotype  $\times$  crop year (GC) interaction, except for FIB (which is determined by a major genetic component). This result confirms a posteriori the interest and relevance of the sites chosen to develop the decentralized selection program likely to exploit or minimize genotype  $\times$  environment interactions. The optimum number of locations likely to sufficiently assess the mean merit over the whole industry of any elite candidate was determined by plotting the trend of

broad-sense heritability ( $H$ ) for each trait for an increasing number of locations. If sugarcane selection were only based on either tonnes of cane per hectare (TCH) or estimable recoverable sugar (ERS), the marginal increase of  $H$  for both of these traits beyond four stations might suggest that not all of them would be necessary. In this hypothetical scenario, resources saved by a reduction in the number of locations could be theoretically guided by considerations relative to the selection of the stations contributing the most to genotype  $\times$  environment interactions (highest ecovariance index) and the elimination of the very few locations most correlated to one of the former. However, the economic index (EI) is the most important trait for selection (along with resistance to diseases), and its evaluation is based on measurement of both TCH and ERS, two traits that are negatively correlated as observed by other authors (Kang et al. 1983; Milligan et al. 1990, 1996; Baffa et al. 2014). As a result, for this key trait, genotypic response appeared either not correlated between locations or slightly correlated between a few of them. No location appeared redundant relative to the others for this EI criterion, in particular for ranking of top elites. This finding confirms and highlights how selection for local adaptation is an important objective, being more desirable than selection for broad adaptation in the context of sugarcane production on Réunion Island. Moreover, the fact that the simulated broad-sense heritability of EI was not firmly capped by a plateau when reaching seven locations fully justifies the objective of seeking additional opportunities for trials to further test the most promising candidate genotypes (if not their complete series) at supplementary locations.

Our study also illustrated that the first plane of GGE biplots can provide reliable summary representations of responses of candidate genotypes across environments and interrelationships among them. This is directly related to the efficiency of the representation of the variation of the initial trait data. The scope of the lessons that can be drawn from this statistical tool needs to be examined for each trait. In our case study, GGE biplots provide at a glance simple and relatively reliable visualizations of both the performance and stability of genotypes across our MET network when compared with the analytical results. Our case study supports the usefulness of this graphical statistical tool to interpret data results and assist decision-making for

selection, further motivating its routine use in sugarcane selection programs.

**Acknowledgements** The authors wish to thank technical staff members J.P. Diorflar, A. Jacquemart, F. Poiny, and P. Langlois for experimental fieldwork, as well as J.M. Sinama, J. Bernard, and C. Roussel for quality analyses of sugarcane. This study was funded by eRcane Company and by Conseil Régional de la Réunion and European Union (European Regional Development Fund—ERDF) through the grant attributed to Valbiocan project.

**Author contributions** T.D. and S.G. carried out the statistical analyses and contributed to writing the paper. A.T.-C. created the experimental populations tested in the field. L.B. supervised the management of the experiment network and the phenotypic database. J.-Y.H. conceived strategies for analyses, contributed to their implementation, and wrote the paper. All authors discussed the results at all stages of the work.

### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflicts of interest.

**Ethical approval** The authors declare that the experiments presented in this publication comply with current French laws.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

### References

- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19:716–723
- Baffa DCF, Costa PMA, Silveira G, Lopes FJF et al (2014) Path analysis for selection of saccharification-efficient sugarcane genotypes through agronomic traits. *Agron J* 106(5):1643–1650
- eRcane (2009) Developing sugar cane resources. 2007–2009 bi-annual report, 75 pp
- Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman, Burnt Mill
- Glaz B, Kang MS (2008) Location contributions determined via GGE biplot analysis of multi environment sugarcane genotype-performance trials. *Crop Sci* 48:941–950
- Hoarau M (1969) Sugar cane analysis by hydraulic press method. *Int Sugar J* 71:328–332
- Holland JB, Nyquist WE, Cervantes-Martinez T (2003) Estimating and interpreting heritability for plant breeding: an update. *Plant Breed Rev* 22:9–112
- Hugot E (1958) Critère et formule de comparaison entre champs, parcelles, variétés ou traitement de cannes. *Rev Agric Sucr l'Ile Maurice* 37:212–216
- Jackson PA, Hogarth DM (1992) Genotype  $\times$  environment interactions in sugarcane: I. patterns of response across sites and crop-years in North Queensland. *Aust J Agric Res* 43(7):1447–1459
- Kang MS, Miller JD, Tai PYP (1983) Genetic and phenotypic path analyses and heritability in sugarcane. *Crop Sci* 23:643–647
- Klomsa-ard P, Jaisil P, Patanothai A (2013) Performance and stability for yield and component traits of elite sugarcane genotypes across production environments in Thailand. *Sugar Tech* 15(4):354–364
- Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O (2006) SAS for mixed models, 2nd edn. SAS Institute Inc., Cary
- Luo J, Pan YB, Xu L, Grisham MP, Zhang H, Zhang H, Que Y (2015) Rational regional distribution of sugarcane cultivars in China. *Sci Rep* 5:15721
- Mariotti JA, Clariana R (1994) The nature of  $G \times E$  interactions in sugarcane hybrid progenies. *Sugar Cane* 3:12–16
- Milligan SB, Gravois KA, Bischoff KP, Martin FA (1990) Crop effects on genetic relationships among sugarcane traits. *Crop Sci* 30:927–931
- Milligan SB, Gravois KA, Martin FA (1996) Inheritance of sugarcane ratooning ability and the relationship of younger crop traits to older crop traits. *Crop Sci* 36:45–50
- Mirzawan PDN, Cooper M, DeLacy IH, Hogarth DM (1994) Retrospective analysis of the relationships among the test environments of the Southern Queensland sugarcane breeding programme. *Theor Appl Genet* 88:707–716
- Parfitt RC (2000) Genotype  $\times$  environment interaction among secondary variety trials in the Northern region of the South African sugar industry. *Proc S Afr Sugarcane Technol Assess* 74:245–248
- Ramburan S, Zhou M, Labuschagne M (2012a) Integrating empirical and analytical approaches to investigate genotype  $\times$  environment interactions in sugarcane. *Crop Sci* 52:2153–2165
- Ramburan S, Zhou M, Labuschagne MT (2012b) Investigating test site similarity, trait relations and causes of genotype  $\times$  environment interactions of sugarcane in the Midlands region of South Africa. *Field Crops Res* 129:71–80
- Rea R, De Souza Vieira O (2002) Genotype  $\times$  environment interactions in sugarcane yield trials in the Central-Western region of Venezuela. *Interciencia* 27(11):620–624
- Saranin AP (1986) The press method of sugarcane analysis. *Proc Aust Soc Sugarcane Technol* 8:33–42
- Shandu SK, Ram B, Kumar S, Singh RS, Uppal SK, Brar KS, Singh P, Nair NV (2012) GGE biplot analysis for visualization of performance rank and stability of sugarcane genotypes. *Int Sugar J* 114:32–41
- Spiegel MR, Lipschutz S, Spellman D (2009) Vector analysis. Schaums outlines, 2nd edn. McGraw Hill, New York
- Wricke G (1965) Zur Berechnung der Ökvalenz bei Sommerweizen und Hafer. *Pflanzenzüchtung* 52:127–138
- Yan W (2002) Singular-value partitioning in biplot analysis of multi-environment trial data. *Agron J* 94:990–996

- Yan W, Tinker NA (2005) An integrated biplot analysis system for displaying, interpreting, and exploring genotype  $\times$  environment interaction. *Crop Sci* 45:1004–1106
- Yan W, Tinker NA (2006) Biplot analysis of multi-environment trial data: principles and applications. *Can J Plant Sci* 86(3):623–645
- Yan W, Hunt LA, Sheng Q, Szlavniks Z (2000) Cultivar evaluation and mega-environment investigation based on GGE biplot. *Crop Sci* 40:597–605